

VU Research Portal

Functional specification of time series models

Bierens, H.J.

1988

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Bierens, H. J. (1988). *Functional specification of time series models*. (Serie Research Memoranda; No. 1988-63). Faculty of Economics and Business Administration, Vrije Universiteit Amsterdam.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

05348

SERIE RESEARCH MEMORANDA

FUNCTIONAL SPECIFICATION OF TIME SERIES MODELS

Herman J. Bierens

Research Memorandum 1988-63

Dec. '88



VRIJE UNIVERSITEIT
FACULTEIT DER ECONOMISCHE WETENSCHAPPEN
EN ECONOMETRIE
AMSTERDAM

FUNCTIONAL SPECIFICATION OF TIME SERIES MODELS *

Herman J. Bierens
Department of Econometrics
Free University, Amsterdam

December 1988

* Chapter 10 of:

PRINCIPLES OF NONLINEAR AND NONPARAMETRIC REGRESSION ANALYSIS

To be published by Marcel Dekker Inc., New York.



TABLE OF CONTENTS:

PREFACE

PART 1: PRELIMINARY MATHEMATICS

1. BASIC PROBABILITY THEORY
 - 1.1 Measure-theoretical foundation of probability theory
 - 1.2 Independence
 - 1.3 Borel measurable functions
 - 1.4 Mathematical expectation
 - 1.5 Characteristic functions
 - 1.6 Random functions
2. CONVERGENCE
 - 2.1 Weak and strong convergence of random variables
 - 2.2 Convergence of mathematical expectations
 - 2.3 Convergence of distributions
 - 2.4 Central limit theorems
 - 2.5 Further results on convergence of distributions and mathematical expectations, and laws of large numbers
 - 2.6 Convergence of random functions
 - 2.7 Uniform strong and weak laws of large numbers
3. INTRODUCTION TO CONDITIONING
 - 3.1 Definition of conditional expectation
 - 3.2. Basic properties of conditional expectations
 - 3.3 Identification of conditional expectations

PART 2: REGRESSION ANALYSIS UNDER INDEPENDENCE

4. NONLINEAR PARAMETRIC REGRESSION ANALYSIS
 - 4.1 Nonlinear regression models and the nonlinear least squares estimator
 - 4.2 Consistency and asymptotic normality: General theory
 - 4.2.1 Consistency
 - 4.2.2 Asymptotic normality
 - 4.3 Consistency and asymptotic normality of nonlinear least squares estimators in the i.i.d. case
 - 4.3.1 Consistency

- 4.3.2 Asymptotic normality
 - 4.3.3 Consistent estimation of the asymptotic variance matrix
- 4.4 Consistency and asymptotic normality of the nonlinear least squares estimator under data heterogeneity
 - 4.4.1 Data heterogeneity
 - 4.4.2 Strong and weak consistency
 - 4.4.3 Asymptotic normality
- 4.5 Testing parameter restrictions: The Wald test
5. TESTS FOR MODEL MISSPECIFICATION
 - 5.1 White's version of Hausman's test
 - 5.2 Newey's M-test
 - 5.2.1 Introduction
 - 5.2.2 The conditional M-test
 - 5.3 A consistent conditional M-test
 - 5.4 The integrated M-test
6. THE NADARAYA-WATSON KERNEL REGRESSION FUNCTION ESTIMATOR
 - 6.1 Introduction
 - 6.2 Asymptotic normality in the continuous case
 - 6.3 Uniform consistency in the continuous case
 - 6.4 Discrete and mixed continuous-discrete regressors
 - 6.4.1 The discrete case
 - 6.4.2 The mixed continuous-discrete case
 - 6.5 The choice of the kernel
 - 6.6 The choice of the window width
 - 6.7 An empirical application to specification of household expenditure systems and equivalence scales
 - 6.7.1 Introduction
 - 6.7.2 Model and data
 - 6.7.3 The results
 - 6.7.4 Sample selection
7. SAMPLE MOMENTS INTEGRATING NORMAL KERNEL ESTIMATORS
 - 7.1 Kernel density estimators
 - 7.1.1 Integral conditions
 - 7.1.2 Sample moments integrating kernel density estimators
 - 7.1.3 The Dirac-catastrophe and the modified SMINK density estimator

- 7.1.4 How to choose the window width of the modified SMINK density estimator
- 7.2 Regression
 - 7.2.1 SMINK estimators of a regression function
 - 7.2.2 How to choose the window width parameters
- 7.3 A numerical example
- 7.4 Proofs
- 8. NONLINEAR REGRESSION WITH DISCRETE EXPLANATORY VARIABLES
 - 8.1. The earnings function
 - 8.2. The functional form of a regression model with discrete explanatory variables
 - 8.3. The choice of the linear separator
 - 8.4. Estimating and testing the regression function
 - 8.4.1 Estimation
 - 8.4.2 Model specification testing
 - 8.4.3 The selection of the polynomial order
 - 8.5 Proofs
 - 8.5.1 Proof of theorem 8.4.2
 - 8.5.2 Proof of theorem 8.4.3

PART 3: TIME SERIES

- 9. CONDITIONING AND DEPENDENCE
 - 9.1 Conditional expectations relative to a Borel field
 - 9.1.1 Definition and basic properties
 - 9.1.2 Martingales
 - 9.1.3 Martingale convergence theorems
 - 9.1.4 A martingale difference central limit theorem
 - 9.2 Measures of dependence
 - 9.2.1 Mixingales
 - 9.2.2 Uniform and strong mixing
 - 9.2.3 ν -Stability
 - 9.3 Weak laws of large numbers for dependent random variables
 - 9.4 Proper heterogeneity and uniform laws for functions of infinitely many random variables
- 10. FUNCTIONAL SPECIFICATION OF TIME SERIES MODELS
 - 10.1 Linear times series regression models
 - 10.1.1 Introduction

- 10.1.2 The Wold decomposition
- 10.1.3 Linear vector time series models
- 10.2 ARMA memory index models
 - 10.2.1 Introduction
 - 10.2.2 Finite conditioning of univariate rational-valued time series
 - 10.2.3 Infinite conditioning of univariate rational-valued time series
 - 10.2.3 The multivariate case
 - 10.2.4 The nature of the ARMA memory index parameters and the response functions
 - 10.2.5 Discussion
- 10.3 Nonlinear ARMAX models
- 11. ARMAX MODELS: ESTIMATION AND TESTING
 - 11.1 Estimation of linear ARMAX models
 - 11.1.1 Introduction
 - 11.1.2 Consistency and asymptotic normality
 - 11.2 Estimation of nonlinear ARMAX models
 - 11.3 A consistent $N(0,1)$ model specification test
 - 11.4 A consistent Hausman-type model specification test
 - 11.5 An autocorrelation test
- 12. NONPARAMETRIC TIME SERIES REGRESSION
 - 12.1 Assumptions and preliminary lemmas
 - 12.2 Consistency and asymptotic normality of time series kernel regression function estimators

PREFACE

This book deals with statistical inference of nonlinear regression models from two opposite points of view, namely the case where the functional form of the model is completely specified as a known function of regressors and unknown parameters, and the opposite case where the functional form of the model is completely unknown. First it is assumed that the response function of the regression model under review belongs to a certain well-specified parametric family of functional forms, by which estimation of the model merely amounts to estimation of the unknown parameters. For this class of models we review the asymptotic properties of the nonlinear least squares estimator for independent data as well as for time series.

In practice assumptions on the functional form are often made on the basis of computational convenience rather than on the basis of precise a priori knowledge of the empirical phenomenon under review. Therefore the linear regression model is still the most popular model specification in applied research. However, even if the specification of the functional form is based on sound theoretical considerations there is quite often a large range of functional forms that are theoretically admissible, so that there is no guarantee that the actually chosen functional form is true. Functional specification of a parametric nonlinear regression model should therefore always be verified by conducting model misspecification tests. Various model misspecification tests will therefore be discussed, in particular consistent tests which have asymptotic power 1 against all deviations from the null hypothesis that the model is correct.

The opposite case of parametric regression is nonparametric regression. Nonparametric regression analysis is concerned with estimation of a regression model without specifying in advance its functional form. Thus the only source of information about the functional form of the model is the data set itself. In this book we shall review various nonparametric regression approaches, with special emphasis on the kernel method, under various distributional assumptions.

This book is divided into three parts. In the first part we review the elements of abstract probability theory we need in part 2. Part 2 is devoted to the asymptotic theory of para-

metric and nonparametric regression analysis in the case of independent data generating processes. In part 3 we extend the analysis involved to time series.

The selection of the topics mainly reflexes my own interest in the subject. Instead of providing an encyclopedic survey of the literature, I have chosen for a setup which aims to fill the gap between intermediate statistics (including linear time series analysis) and the level necessary to get access to the recent literature on nonlinear and nonparametric regression analysis, with emphasis on my own contributions. The ultimate goal is to provide the student with the tools for his own independent research in this area, by showing what tools I and others have used and what they have been used for. Thus, this book may be viewed as an account of my own struggle with the material involved. I think this book is particularly suitable for self-tuition (at least it aims to be), and may prove useful in a graduate course in mathematical statistics and advanced econometrics.

Acknowledgements:

The first five chapters of this book have been disseminated in draft form as working papers. I am grateful to Anil Bera, Alexander Georgiev and Jan Magnus for suggesting additional references, and in particular to Lourens Broersma, Johan Smits and Ton Steerneman who suggested various improvements.

A large body of the material in chapter 6 has been published earlier in Truman F. Bewley (ed.), *Advances in Econometrics, Fifth World Congress*, Cambridge University Press. I am indebted to Cambridge University Press for granting permission to reprint it.

10. FUNCTIONAL SPECIFICATION OF TIME SERIES MODELS

10.1 Introduction

Consider a vector time series process (Z_t) in R^k with $E|Z_t| < \infty$ for each t . In time series regression analysis we are interested in modeling and estimating the conditional expectation of Z_t relative to its entire past. The reason for our interest in this conditional expectation is that it represents the best forecasting scheme for Z_t ; best in the sense that the mean square forecast error is minimal. To see this, compare this forecast, i.e.,

$$\hat{Z}_t = E(Z_t | Z_{t-1}, Z_{t-2}, \dots), \quad (10.1.1)$$

with an alternative forecasting scheme, say

$$\bar{Z}_t = G_t(Z_{t-1}, Z_{t-2}, \dots), \quad (10.1.2)$$

where G_t is a R^k -valued (non-random) function on the space of all one-sided infinite sequences in R^k such that \bar{Z}_t is a well-defined random vector. Denote the forecast errors by

$$U_t = Z_t - \hat{Z}_t, \quad (10.1.3)$$

and

$$W_t = Z_t - \bar{Z}_t, \quad (10.1.4)$$

respectively. Then

$$E W_t W_t' = E U_t U_t' + E(\hat{Z}_t - \bar{Z}_t)(\hat{Z}_t - \bar{Z}_t)', \quad (10.1.5)$$

due to the fact that by (10.1.1) and (10.1.3)

$$E(U_t | Z_{t-1}, Z_{t-2}, \dots) = 0 \quad \text{a.s.} \quad (10.1.6)$$

Thus the mean square error matrix $E W_t W_t'$ of the alternative forecasting scheme dominates $E U_t U_t'$ by a positive semidefinite matrix.

10.2 Linear time series regression models

10.2.1 The Wold decomposition

The central problem in time series modeling is to find a suitable functional specification of the conditional expectation (10.1.1). Often the model is specified directly or indirectly as a linear AR(∞) model. This linear AR(∞) specification can be motivated on the basis of the famous Wold decomposition theorem, together with the assumption that the process (Z_t) is stationary and Gaussian, and some regularity conditions. Here we give a special case of Wold's theorem, for univariate time series processes.

Theorem 10.2.1 (Wold decomposition). Let (Z_t) be a univariate stationary Gaussian time series process satisfying $E Z_t = 0$. Let

$$\sigma^2 = E U_t^2 > 0,$$

where U_t is defined by (10.1.3), and let for $s \geq 0$,

$$\gamma_s = E Z_t U_{t-s} / \sigma^2 \quad (\text{Note that } \gamma_0 = 1).$$

Then

$$Z_t = \sum_{s=0}^{\infty} \gamma_s U_{t-s} + W_t$$

where the process (W_t) is such that

$$E W_j U_t = 0 \quad \text{for all } j \text{ and } t;$$

(W_t) is deterministic,

i.e., W_t is a (possibly infinite) linear combination of W_{t-1} , W_{t-2} , ... without error. Moreover,

$$\sum_{s=0}^{\infty} \gamma_s^2 < \infty;$$

(U_t) is an independent Gaussian process. (10.2.1)

Proof: Since Z_t is stationary and Gaussian $E(Z_t | Z_{t-1}, \dots, Z_{t-m})$ is a linear function of Z_{t-1}, \dots, Z_{t-m} , i.e., there exist constants $\beta_{1,m}, \dots, \beta_{m,m}$ such that

$$E(Z_t | Z_{t-1}, \dots, Z_{t-m}) = \sum_{j=1}^m \beta_{j,m} Z_{t-j} \quad (10.2.2)$$

(Cf. exercise 1). Defining

$$U_{t,m} = Z_t - E(Z_t | Z_{t-1}, \dots, Z_{t-m})$$

it follows that $(U_{t,m}, Z_{t-1}, \dots, Z_{t-m})$ is $(m+1)$ -variate normally distributed with $U_{t,m}$ independent of Z_{t-1}, \dots, Z_{t-m} . Since by theorem 9.1.4,

$$\lim_{m \rightarrow \infty} U_{t,m} = U_t \quad \text{a.s.} \quad (10.2.3)$$

it follows that U_t is independent of $(Z_{t-1}, \dots, Z_{t-\ell})$ for every $\ell \geq 1$, $U_t \sim N(0, \sigma^2)$ and U_t is a linear combination of $Z_t, Z_{t-1}, Z_{t-2}, \dots$. With these hints, (10.2.1) is not hard to prove (Cf. exercise 2). The rest of the proof now follows from the original Wold decomposition theorem. See, e.g., Anderson (1971, pp.420-421). Q.E.D.

The next step is to assume that

$$\text{the deterministic process } (W_t) \text{ is zero} \quad (10.2.4)$$

and that the lag polynomial

$$\gamma(L) = \sum_{s=0}^{\infty} \gamma_s L^s$$

is invertible:

$$\beta(L) = \gamma(L)^{-1} = \sum_{s=0}^{\infty} \beta_s L^s. \quad (10.2.5)$$

See Anderson (1971, pp 423-424) for precise conditions under which (10.2.5) holds. Then

$$\sum_{s=0}^{\infty} \beta_s Z_{t-s} = U_t, \quad (10.2.6)$$

hence, since $\beta_0=1$,

$$Z_t = \sum_{s=1}^{\infty} (-\beta_s) Z_{t-s} + U_t, \quad (10.2.7)$$

which is an AR(∞) model.

In practice one often assumes that the lag polynomial $\gamma(L)$ is rational, i.e.

$$\gamma(L) = \theta(L)/\alpha(L), \quad (10.2.8)$$

where

$$\theta(L) = 1 - \sum_{s=1}^q \theta_s L^s \quad (10.2.9)$$

$$\alpha(L) = 1 - \sum_{s=1}^p \alpha_s L^s \quad (10.2.10)$$

are finite-order lag polynomials with no common roots, and all roots outside the unit circle. Then, with condition (10.2.4),

$$\alpha(L)Z_t = \theta(L)U_t, \quad (10.2.11)$$

which is an ARMA(p,q) model.

10.2.2 Linear vector time series models

Similar results also hold for vector time series processes. If (Z_t) is a k-variate stationary Gaussian process then under some regularity conditions we have

$$Z_t = \sum_{s=0}^{\infty} \Gamma_s U_{t-s} \quad (10.2.12)$$

where $\Gamma_s = (E Z_t U_{t-s}') (E U_t U_t')^{-1}$. Again assuming that the matrix-valued lag polynomial

$$\Gamma(L) = \sum_{s=0}^{\infty} \Gamma_s L^s \quad (\Gamma_0 = I) \quad (10.2.13)$$

is invertible with inverse

$$B(L) = \Gamma(L)^{-1} = \sum_{s=0}^{\infty} B_s L^s \quad (B_0 = I) \quad (10.2.14)$$

the model becomes a VAR(∞) model:

$$Z_t = \sum_{s=1}^{\infty} (-B_s) Z_{t-s} + U_t. \quad (10.2.15)$$

If $\Gamma(L)$ is rational, i.e.

$$\Gamma(L) = A(L)^{-1}\Theta(L) \quad (10.2.16)$$

with

$$A(L) = I - \sum_{s=1}^p A_s L^s \quad (10.2.17)$$

$$\Theta(L) = I - \sum_{s=1}^q \Theta_s L^s, \quad (10.2.18)$$

then (under some regularity conditions), the model becomes a VARMA(p,q) model:

$$A(L)Z_t = \Theta(L)U_t \quad (10.2.19)$$

with $A(0) = \Theta(0) = I$.

VARMA models, however, may be considered as systems of ARMAX models. This is obvious if $\Theta(L)$ is diagonal, but also if not each equation in (10.2.19) can be written as an ARMAX model. To see this, observe that

$$\Theta(L)^{-1} = (\det \Theta(L))^{-1}C(L),$$

where $C(L)$ is the matrix of co-factors of $\Theta(L)$. Multiplying both sides of (10.2.19) by $C(L)$ then yields

$$C(L)A(L)Z_t = (\det \Theta(L))U_t. \quad (10.2.20)$$

The polynomial matrix $\Psi(L) = C(L)A(L)$ consists of finite-order lag polynomials, and also $\varphi(L) = \det \Theta(L)$ is a finite-order lag polynomial. The first equation of (10.2.20) therefore takes the form

$$Z_{1,t} + \sum_{i=1}^k \sum_{j=1}^{p^*} \psi_{1,i,j} Z_{1,t-j} = U_{1,t} + \sum_{j=1}^{q^*} \gamma_j U_{1,t-j} \quad (10.2.21)$$

Denoting

$$Y_t = Z_{1,t}, \quad X_t = (Z_{2,t}, \dots, Z_{k,t})', \quad V_t = U_{1,t}$$

$$\alpha_j = -\psi_{1,1,j}, \quad \beta_j = (-\psi_{1,2,j}, \dots, -\psi_{1,k,j})',$$

we can now write (10.2.21) in ARMAX form as

$$Y_t = \sum_{j=1}^{p^*} \alpha_j Y_{t-j} + \sum_{j=1}^{p^*} \beta_j' X_{t-j} + V_t + \sum_{j=1}^{q^*} \gamma_j V_{t-j} \quad (10.2.22)$$

Finally, for the case $q^*=0$ we get an ARX-model.

Exercises:

2. Prove (10.2.2).
3. Prove (10.2.1).

10.3 ARMA memory index models.

10.3.1 Introduction

The linearity of the time series models discussed in section 10.2 is due to the assumption of normality of the time series involved. Normality, however, is by no means a necessity for time series. So the question arises what can be said about the functional form of the conditional expectation (10.1.1) if the process (Z_t) is non-Gaussian.

In this section we discuss the ARMA memory index modeling approach of Bierens (1988a,b). This approach exploits the fact that all time series are rational-valued. One could consider the rationality condition as an assumption, but in practice one cannot deal with irrational numbers, hence time series are always reported in a finite number of decimal digits and consequently time series are rational-valued by nature. Thus, the rationality condition is an indisputable fact rather than an assumption.

In this section it will be shown that in conditioning a k -variate rational-valued time series process on its entire past it is possible to capture the information contained in the past of the process by a single random variable. This random variable, containing all relevant information about the past of the process involved, can be formed as an autoregressive moving average of past observations. Hence the conditional expectation involved then takes the form of a nonlinear function of an autoregressive moving average of past observations. In particular, for univariate rational-valued time series processes (Z_t) it will be shown that there exist uncountably many real numbers $\tau \in (-1,1)$ such that

$$E(Z_t | Z_{t-1}, Z_{t-2}, Z_{t-3}, \dots) = E(Z_t | \sum_{j=1}^{\infty} r^{j-1} Z_{t-j}) \text{ a.s.} \quad (10.3.1)$$

Moreover, if Z_t is k -variate rational-valued there exist uncountably many $r \in (-1, 1)$ and $\theta \in \mathbb{R}^k$ such that for $i=1, \dots, k$,

$$E(Z_{i,t} | Z_{t-1}, Z_{t-2}, Z_{t-3}, \dots) = E(Z_{i,t} | \sum_{j=1}^{\infty} r^{j-1} \theta' Z_{t-j}) \text{ a.s.} \quad (10.3.2)$$

where $Z_{i,t}$ is the i -th component of Z_t .

This result is not specific for the geometric weighting scheme involved. More generally, it will be shown that there exist uncountably many sets of rational lag polynomials

$$\psi_{i,j}(L) = \psi_{i,j}^{(1)}(L) / \psi_i^{(2)}(L), \quad (i, j = 1, \dots, k),$$

where

$$\psi_{i,j}^{(1)}(L) \text{ and } \psi_i^{(2)}(L)$$

are finite-order lag polynomials, such that for $i=1, 2, \dots, k$,

$$E(Z_{i,t} | Z_{t-1}, Z_{t-2}, Z_{t-3}, \dots) = E(Z_{i,t} | \sum_{j=1}^k \psi_{i,j}(L) Z_{j,t-1}) \text{ a.s.} \quad (10.3.3)$$

Since a conditional expectation can be written as a Borel measurable function of the conditioning variable, the result (10.3.3) implies that for each permissible lag polynomial $\psi_{i,j}(L)$ there exists a Borel measurable real function $f_{i,t}$ such that

$$E(Z_{i,t} | Z_{t-1}, Z_{t-2}, Z_{t-3}, \dots) = f_{i,t}(\sum_{j=1}^k \psi_{i,j}(L) Z_{j,t-1}) \text{ a.s.} \quad (10.3.4)$$

Denoting

$$\xi_{i,t} = \sum_{j=1}^k \psi_{i,j}(L) Z_{j,t-1} = \sum_{j=1}^k \{\psi_{i,j}^{(1)}(L) / \psi_i^{(2)}(L)\} Z_{j,t-1}, \quad (10.3.5)$$

we see that the conditioning variable in (10.3.3) can be

written in ARMA form:

$$\psi_i^{(2)}(L)\xi_{i,t} = \sum_{j=1}^k \psi_{i,j}^{(1)}(L)Z_{j,t-1}. \quad (10.3.6)$$

Consequently, specifying the data generating process as an ARMA process is equivalent to specifying the response functions $f_{i,t}$, for a particular set of rational lag polynomials $\psi_{i,j}(L)$, as time invariant linear functions. Moreover, in the multivariate case one may interpret the conditioning variable $\xi_{i,t}$ as a one-step ahead forecast with an almost arbitrary linear ARMAX model for $Z_{i,t}$. This can be seen if one replaces $\xi_{i,t}$ in (10.3.6) by $Z_{i,t} - V_{i,t}$, where $(V_{i,t})$ is the error process. The X-vector involved then consists of all components of Z_t except $Z_{i,t}$. Thus, the best one-step ahead forecasting scheme is a Borel measurable real function of a one-step ahead forecast with an almost arbitrary linear ARMAX model. Specifying the equations in the VARMA model (10.2.19) as ARMAX models is therefore equivalent to specifying the corresponding functions $f_{i,t}$ in (10.3.4) as linear time invariant functions. Furthermore, all the non-linearity of the conditional expectation function (10.3.4) is now captured by the nonlinearity of the functions $f_{i,t}$, and the impact of heterogeneity of the process (Z_t) on the conditional expectation involved is captured by the time dependence of $f_{i,t}$.

As the conditioning variable (10.3.5) carries the memory of the process, plays a similar role as the index in the index modeling approach of Sargent and Sims (1977) and Sims (1981), and can be written in ARMA(X) form, we have called our approach *Auto-Regressive Moving Average (ARMA) Memory Index Modeling* and the index (10.3.5) will be called the *ARMA memory index*.

10.3.2 Finite conditioning of univariate rational-valued time series

Let (Z_t) be a Q -valued stochastic process, where Q is the set of rational numbers. If

$$E|Z_t| < \infty \text{ for every } t, \quad (10.3.7)$$

then $E(Z_t | Z_{t-1}, \dots, Z_{t-m})$ exists for any integers t and $m \geq 1$ (see chapter 3). Now our aim is to show that the conditioning variables Z_{t-1}, \dots, Z_{t-m} in this conditional expectation may be replaced by $\sum_{j=1}^m r^{j-1} Z_{t-j}$ for some real numbers r , provided m is finite.

Suppose there exists a Borel measurable one-to-one mapping Φ_m from Q^m (the m -dimensional space of vectors with rational-valued components) to a subset of R . Then

$$(Z_{t-1}, \dots, Z_{t-m}) \text{ and } \Phi_m(Z_{t-1}, \dots, Z_{t-m})$$

generate the same Borel field, hence by the definition of conditional expectation

$$E(Z_t | Z_{t-1}, \dots, Z_{t-m}) = E(Z_t | \Phi_m(Z_{t-1}, \dots, Z_{t-m})) \quad (10.3.8)$$

for each t , provided (10.3.7) holds. Thus we see that by using such a one-to-one mapping Φ_m we can reduce the number of conditioning variables from m to one. This easy result, which is reminiscent of the approach in chapter 8, is the core of our approach.

The function Φ_m may be constructed as follows. Let

$$\Phi_m(w|r) = \sum_{j=1}^m w_j r^{j-1}, \quad (10.3.9)$$

where

$$w = (w_1, \dots, w_m)' \in Q^m, \quad r \in R. \quad (10.3.10)$$

Moreover, let for $w^{(1)} \in Q^m, w^{(2)} \in Q^m$,

$$R_m(w^{(1)}, w^{(2)}) = \{r \in R : \Phi_m(w^{(1)}|r) = \Phi_m(w^{(2)}|r)\}. \quad (10.3.11)$$

In other words, $R_m(w^{(1)}, w^{(2)})$ is the set of real roots of the $(m-1)$ -order polynomial

$$\Phi_m(w^{(1)} | r) - \Phi_m(w^{(2)} | r) = \sum_{j=1}^m (w_j^{(1)} - w_j^{(2)}) r^{j-1}. \quad (10.3.12)$$

It is well-known that if $w^{(1)} \neq w^{(2)}$, so that for at least one j ,

$$w_j^{(1)} \neq w_j^{(2)},$$

the number of real roots of the polynomial (10.3.12) does not exceed $m-1$. Thus if $w^{(1)} \neq w^{(2)}$ then $R_m(w^{(1)}, w^{(2)})$ is a finite set of size less than or equal to $m-1$. Since Q is a countable set [see Royden (1968, proposition 6 at p.21)] and since the union of a countable collection of countable sets is countable itself [Royden (1968, proposition 7 at p.21)], we now obviously have that the set

$$S_m = \cup R_m(w^{(1)}, w^{(2)}) \quad (10.3.13)$$

is countable, where the union is over all Q^m -valued unequal $w^{(1)}$ and $w^{(2)}$. Thus for $r \in R \setminus S_m$ we have that

$$w^{(1)} \in Q^m, w^{(2)} \in Q^m, \Phi_m(w^{(1)} | r) = \Phi_m(w^{(2)} | r) \Rightarrow w^{(1)} = w^{(2)} \quad (10.3.14)$$

and vice versa. This proves that for $r \in R \setminus S_m$ the function $\Phi_m(w | r)$ is a one-to-one mapping from Q^m to a subset of R . Taking

$$S = \cup_{m=1}^{\infty} S_m, \quad (10.3.15)$$

which is a countable union of countable sets and therefore countable itself, we now see that the following theorem holds.

Theorem 10.3.1 Let (Z_t) be a Q -valued stochastic process satisfying $E|Z_t| < \infty$ for all t . Then there exists a countable subset S of R such that $r \in R \setminus S$ implies

$$E(Z_t | Z_{t-1}, \dots, Z_{t-m}) = E(Z_t | \sum_{j=1}^m Z_{t-j} r^{j-1}) \text{ a.s.} \quad (10.3.16)$$

for $m=1, 2, 3, \dots$ and $t=\dots, -2, -1, 0, 1, 2, 3, \dots$

Remark: Note that this result carries over for processes (Z_t) in any countable subset of R , as we only have used the countability of Q for proving Theorem 10.3.1. Thus, the theorem remains valid if the Z_t are Borel measurable transformations of Q -valued random variables, for countability is always preserved.

10.3.3 Infinite conditioning of univariate rational-valued time series

In this subsection we shall set forth conditions such that (10.3.1) holds for each t and each $r \in (-1,1) \setminus S$, where S is the same as in theorem 10.3.1. Intuitively we feel that (10.3.1) requires the following condition:

The process (Z_t) is such that for every t and every

$$r \in (-1,1), \sum_{j=1}^{\infty} Z_{t-j} r^{j-1} \text{ converges a.s.} \quad (10.3.17)$$

As has been shown in Bierens (1988a), this condition is implied by the following assumption:

Assumption 10.3.1 Let $\sup_t E|Z_t| < \infty$.

Now if condition (10.3.17) holds then

$$(Z_{t-1}, Z_{t-2}, \dots) \text{ and } (\sum_{j=1}^{\infty} Z_{t-j} r^{j-1}, Z_{t-m-1}, Z_{t-m-2}, \dots)$$

generate the same Borel field, because both sequences can then be constructed from

$$(\sum_{j=1}^m Z_{t-j} r^{j-1}, Z_{t-m-1}, Z_{t-m-2}, \dots)$$

and vice versa. Since this conclusion holds for every $m \geq 1$, it follows that under the conditions of theorem 10.3.1 and condition (10.3.17) or assumption 10.3.1, $r \in (-1,1) \setminus S$ implies

$$E(Z_t | Z_{t-1}, Z_{t-2}, \dots) = E(Z_t | \sum_{j=1}^{\infty} Z_{t-j} r^{j-1}, Z_{t-m-1}, Z_{t-m-2}, \dots) \text{ a.s.}$$

for every $m \geq 1$ and t , hence for every t ,

$$\begin{aligned}
E(Z_t | Z_{t-1}, Z_{t-2}, \dots) &= \\
&= \lim_{m \rightarrow \infty} E(Z_t | \sum_{j=1}^{\infty} Z_{t-j} \tau^{j-1}, Z_{t-m-1}, Z_{t-m-2}, \dots) \text{ a.s.}
\end{aligned}
\tag{10.3.18}$$

For showing (10.3.1) we now need additional conditions ensuring that the impact of $Z_{t-m-1}, Z_{t-m-2}, \dots$ on the conditional expectation at the right hand side of (10.3.18) vanishes as $m \rightarrow \infty$. In Bierens (1988a,b) we have shown that ν -stability with respect to an α -mixing base, together with some regularity conditions, will do. However, the proof involved is rather complicated. Therefore we impose here the following extension of the mixingale condition.

Assumption 10.3.2 Let F_{∞}^t be the Borel field generated by

$$Z_t, Z_{t-1}, Z_{t-2}, Z_{t-3}, \dots$$

and let F_t^{∞} be the Borel field generated by

$$Z_t, Z_{t+1}, Z_{t+2}, Z_{t+3}, \dots$$

Let $\ell < \infty$ be an arbitrary integer and let W_t be an arbitrary random variable defined on $F_{t-\ell}^{\infty}$ satisfying $E W_t^2 < \infty$. Moreover, let $\{G_{\infty}^t\}$ and $\{H_{\infty}^t\}$ be arbitrary sequences of Borel fields such that $G_{\infty}^t \subset F_{\infty}^t$ and $H_{\infty}^t \subset F_{\infty}^t$. For every t and every $m \geq 0$ there exist constants c_t, ψ_m , with $\psi_m \rightarrow 0$ as $m \rightarrow \infty$, such that

$$[E(E(W_t | G_{\infty}^{t-1} \vee H_{\infty}^{t-m-1}) - E(W_t | G_{\infty}^{t-1}))^2]^{1/2} \leq c_t \psi_m.$$

This assumption is stated more generally than needed here. We will need its full extent in chapter 11.

Admittedly, assumption 10.3.2 looks quite complicated. However, it simply states that the impact of the remote past of Z_t , where the remote past involved is represented by H_{∞}^{t-m-1} , vanishes as $m \rightarrow \infty$.

In the case (10.3.18), $W_t = Z_t$, $H_{\infty}^t = F_{\infty}^t$ and G_{∞}^{t-1} is the Borel field generated by $\sum_{j=1}^{\infty} Z_{t-j} \tau^{j-1}$, hence assumption 10.3.2 and Chebishev's inequality imply

$$\begin{aligned} \text{plim}_{m \rightarrow \infty} E(Z_t | \sum_{j=1}^{\infty} Z_{t-j} r^{j-1}, Z_{t-m-1}, Z_{t-m-2}, \dots) \\ = E(Z_t | \sum_{j=1}^{\infty} Z_{t-j} r^{j-1}). \end{aligned}$$

From theorem 2.1.6 it now follows that there exists a subsequence (m_ℓ) such that

$$\begin{aligned} E(Z_t | \sum_{j=1}^{\infty} Z_{t-j} r^{j-1}, Z_{t-m_\ell-1}, Z_{t-m_\ell-2}, \dots) \\ \rightarrow E(Z_t | \sum_{j=1}^{\infty} Z_{t-j} r^{j-1}) \text{ a.s. as } \ell \rightarrow \infty, \end{aligned}$$

hence the limit (10.3.18) must be equal to the latter conditional expectation as well. Thus we have :

Theorem 10.3.2 Let the conditions of theorem 10.3.1 and assumptions 10.3.1 and 10.3.2 be satisfied. There exists a countable subset S of R such that $r \in (-1, 1) \setminus S$ implies

$$E(Z_t | Z_{t-1}, Z_{t-2}, \dots) = E(Z_t | \sum_{j=1}^{\infty} Z_{t-j} r^{j-1}) \text{ a.s.}$$

for $t = \dots, -2, -1, 0, 1, 2, 3, \dots$

10.3.4 The multivariate case

If $Z_t \in Q^k$ we may proceed in the same way as before, hence theorems 10.3.1 and 10.3.2 carry over to rational-valued vector time series processes. However, the ARMA memory index $\sum_{j=1}^{\infty} Z_{t-j} r^{j-1}$ is then multivariate too.

We can get a scalar ARMA memory index by using the concept of a linear separator introduced in chapter 8 (cf. definition 8.2.1). Thus, let $\theta \in R^k$ be a linear separator of the countable set Q^k . Then

$$E(Z_t | Z_{t-1}, Z_{t-2}, \dots) = E(Z_t | \theta' Z_{t-1}, \theta' Z_{t-2}, \dots) \text{ a.s.}$$

and moreover the process $(\theta' Z_t)$ is still countable-valued. Applying theorem 10.3.2 we now conclude that for each linear separator θ of Q^k there exists a countable subset S_θ of R such that for each $r \in (-1, 1) \setminus S_\theta$,

$$E(Z_t | Z_{t-1}, Z_{t-2}, \dots) = E(Z_t | \sum_{j=1}^{\infty} \theta' Z_{t-j} r^{j-1}) \text{ a.s. (10.3.19)}$$

We recall that the set Θ_0 of vectors $\theta \in R^k$ that are not linear separators of Q^k has Lebesgue measure zero. See theorem 8.2.1. This result, together with the countability of S_θ for each linear separator θ , imply that there exists a set $N \subset R^{k+1}$ with Lebesgue measure zero such that (10.3.19) holds for $(\theta, r) \in R^k \times (-1, 1) \setminus N$. To see this, draw r from the uniform $[-1, 1]$ distribution, draw the components of θ independently from say the uniform $[a, b]$ distribution and let

$$\zeta_t(r, \theta) = E(I[E(Z_t | Z_{t-1}, Z_{t-2}, \dots) = E(Z_t | \sum_{j=1}^{\infty} \theta' Z_{t-j} r^{j-1})] | r, \theta)$$

Then

$$\begin{aligned} E \zeta_t(r, \theta) &= \int_{\times_{i=1}^k [a, b]} \left(\int_{[-1, 1]} \zeta_t(r, \theta) dr \right) d\theta \\ &= \int_{(\times_{i=1}^k [a, b]) \setminus \Theta_0} \left(\int_{(-1, 1) \setminus S_\theta} \zeta_t(r, \theta) dr \right) d\theta = 1, \end{aligned}$$

which implies that (10.3.19) holds except for (θ, r) in a set with Lebesgue measure zero. Thus we have:

Theorem 10.3.3. Let $(Z_t) = ((Z_{1,t}, \dots, Z_{k,t})')$ be a Q^k -valued stochastic process satisfying $E|Z_t| < \infty$ for every t . There exists a subset N of R^{k+1} with Lebesgue measure zero such that $(\theta, r) \in R^k \times (-1, 1) \setminus N$ implies

$$E(Z_{i,t} | Z_{t-1}, \dots, Z_{t-m}) = E(Z_{i,t} | \sum_{j=1}^m \theta' Z_{t-j} r^{j-1}) \text{ a.s. (10.3.20)}$$

for $m=1, 2, \dots, i=1, 2, \dots, k$ and $t=\dots, -2, -1, 0, 1, 2, 3, \dots$. If in addition assumptions 10.3.1 and 10.3.2 are satisfied, then

$$E(Z_{i,t} | Z_{t-1}, Z_{t-2}, \dots) = E(Z_{i,t} | \sum_{j=1}^{\infty} \theta' Z_{t-j} r^{j-1}) \text{ a.s. (10.3.21)}$$

for $i=1, 2, \dots, k$ and $t=\dots, -2, -1, 0, 1, 2, 3, \dots$

For showing that (10.3.3) holds for uncountably many rational lag polynomials we need the following generalisation of theorems 10.3.1 and 10.3.2.

Lemma 10.3.1. Let (Z_t) be a Q -valued stochastic process with $E|Z_t| < \infty$. Let q be an arbitrary positive integer. Let C_1 be the set of complex numbers with absolute value less than 1. There exists a subset S of C_1^q with Lebesgue measure zero such that $(\tau_1, \dots, \tau_q)' \in C_1^q \setminus S$ implies

$$E(Z_t | Z_{t-1}, \dots, Z_{t-m_q}) = E(Z_t | \prod_{\ell=1}^q \{(1-(\tau_\ell L)^m)/(1-\tau_\ell L)\} Z_{t-1}) \text{ a.s.} \quad (10.3.22)$$

for $m=1,2,3,\dots$ and $t=\dots,-2,-1,0,1,2,3,\dots$. If in addition assumptions 10.3.1 and 10.3.3 hold then

$$E(Z_t | Z_{t-1}, Z_{t-2}, \dots) = E(Z_t | \prod_{\ell=1}^q (1/(1-\tau_\ell L)) Z_{t-1}) \text{ a.s.} \quad (10.3.23)$$

for $t=\dots,-2,-1,0,1,2,3,\dots$

Proof: Let (w_t) be an arbitrary sequence of rational numbers. Denote for $q \geq 1$, $m \geq 1$,

$$x_t^{(1)}(\tau_1) = \sum_{j=0}^{m-1} \tau_1^j w_{t-j} = ((1-(\tau_1 L)^m)/(1-\tau_1 L)) w_t,$$

$$\begin{aligned} x_t^{(q)}(\tau_1, \tau_2, \dots, \tau_q) &= \sum_{j=0}^{m-1} \tau_q^j x_{t-j}^{(q-1)}(\tau_1, \tau_2, \dots, \tau_{q-1}) \\ &= \prod_{\ell=1}^q ((1-(\tau_\ell L)^m)/(1-\tau_\ell L)) w_t. \end{aligned}$$

Suppose for the moment that τ_1, \dots, τ_q are real-valued. Now draw $\bar{\tau}_1, \dots, \bar{\tau}_q$ independently from the uniform $[0,1]$ distribution. Then

$$P[x_t^{(q)}(\bar{\tau}_1, \dots, \bar{\tau}_q) = 0 | \bar{\tau}_1, \dots, \bar{\tau}_q] = 0 \text{ a.s.}$$

if at least one of the

$$x_{t-j}^{(q-1)}(\bar{\tau}_1, \dots, \bar{\tau}_{q-1}), \quad j=0, \dots, m-1,$$

is unequal to zero, whereas

$$P[x_t^{(q)}(\bar{\tau}_1, \dots, \bar{\tau}_q) = 0 | \bar{\tau}_1, \dots, \bar{\tau}_q] = 1 \text{ a.s.}$$

if all the $x_{t-j}^{(q-1)}(\bar{\tau}_1, \dots, \bar{\tau}_{q-1})$ are zero. Thus

$$\begin{aligned}
P[x_t^{(q)}(\tau_1, \dots, \tau_q) = 0 | \tau_1, \dots, \tau_q] \\
= \min_{j=0, \dots, m-1} I[x_{t-j}^{(q-1)}(\tau_1, \dots, \tau_{q-1}) = 0]
\end{aligned}$$

and consequently

$$\begin{aligned}
P[x_t^{(q)}(\tau_1, \dots, \tau_q) = 0] \\
= E \min_{j=0, \dots, m-1} I[x_{t-j}^{(q-1)}(\tau_1, \dots, \tau_{q-1}) = 0] \\
\leq \min_{j=0, \dots, m-1} P[x_{t-j}^{(q-1)}(\tau_1, \dots, \tau_{q-1}) = 0].
\end{aligned}$$

By recursion it therefore follows

$$\begin{aligned}
P[x_t^{(q)}(\tau_1, \dots, \tau_q) = 0] \\
\leq \min_{j=0, \dots, q(m-1)} P[x_{t-j}^{(1)}(\tau_1) = 0]. \quad (10.3.24)
\end{aligned}$$

But

$$\begin{aligned}
P(x_{t-j}^{(1)}(\tau_1) = 0) &= 0 \text{ if at least one } w_{t-j} \text{ (} j=0, \dots, m-1 \text{)} \\
&\text{is unequal to zero,} \\
P(x_{t-j}^{(1)}(\tau_1) = 0) &= 1 \text{ if } w_{t-j} = 0 \text{ for } j=0, \dots, m-1,
\end{aligned}$$

hence

$$P(x_t^{(1)}(\tau_1) = 0) = \min_{j=0, \dots, m-1} I(w_{t-j} = 0). \quad (10.3.25)$$

Combining (10.3.24) and (10.3.25) now yields

$$P[x_t^{(q)}(\tau_1, \dots, \tau_q) = 0] \leq \min_{j=0, \dots, (m-1)q} I(w_{t-j} = 0).$$

This result shows that there exists a subset S_* of $\times_{j=1}^q (-1, 1)$, depending on $w_t, w_{t-1}, \dots, w_{t-q(m-1)}$, which has Lebesgue measure zero if one of these w_{t-j} 's is unequal to zero. The set S in lemma 10.3.1 is now the countable union of all these null sets S_* .

The case that the τ_ℓ are complex-valued is similar. For example, let

$$\tau_\ell = \bar{\rho}_\ell (\cos \bar{\psi}_\ell + i \sin \bar{\psi}_\ell),$$

where the \bar{p}_ℓ are drawn independently from the uniform $[-1,1]$ distribution and the $\bar{\psi}_\ell$ are drawn independently from say the standard normal distribution.

The rest of the proof of lemma 10.3.1 similar to the proofs of theorems 10.3.1 and 10.3.2. Q.E.D.

Now let Γ_q be the set of vectors $\gamma = (\gamma_1, \dots, \gamma_q)' \in \mathbb{R}^q$ for which the polynomial $1 + \sum_{s=1}^q \gamma_s L^s$ has roots all outside the unit circle. Realizing that these roots are related to $\gamma_1, \dots, \gamma_q$ by a one-to-one mapping, the following corollary of part (10.3.23) of lemma 10.3.1 is easy to verify.

Lemma 10.3.2. Let the conditions of part (10.3.23) of lemma 10.3.1 hold. There exists a subset S of Γ_q with Lebesgue measure zero such that $\gamma = (\gamma_1, \dots, \gamma_q)' \in \Gamma_q \setminus S$ implies

$$E(Z_t | Z_{t-1}, Z_{t-2}, \dots) = E(Z_t | (1/(1 + \sum_{s=1}^q \gamma_s L^s)) Z_{t-1}) \text{ a.s.}$$

for $t = \dots, -2, -1, 0, 1, 2, 3, \dots$

Moreover, part (10.3.21) of theorem 10.3.3 can now be generalized as follows:

Lemma 10.3.3. Let the conditions of part (10.3.21) of theorem 10.3.3 hold and let q be an arbitrary positive integer. There exists a subset N of $\mathbb{R}^k \times \Gamma_q$ with Lebesgue measure zero such that $(\theta, \gamma) \in \mathbb{R}^k \times \Gamma_q \setminus N$ implies

$$E(Z_t | Z_{t-1}, Z_{t-2}, \dots) = E(Z_{i,t} | (1/(1 + \sum_{s=1}^q \gamma_s L^s)) \theta' Z_{t-1}) \text{ a.s.}$$

for $i=1, 2, \dots, k$ and $t = \dots, -2, -1, 0, 1, 2, 3, \dots$

Proof: Similarly to theorem 10.3.3.

Applying lemma 10.3.3 to the sequence:

$$(Z_t^*) \text{ with } Z_t^* = (Z_{1,t}, Z_{2,t}, \dots, Z_{t-p,t})',$$

our main theorem below now easily follows.

Theorem 10.3.4. Let p and q be arbitrary integers satisfying $p \geq 0$, $q \geq 1$. Let

$$\psi(L|\beta, \gamma) = (\sum_{s=0}^p \beta_s L^s) / (1 + \sum_{s=1}^q \gamma_s L^s),$$

where $\beta = (\beta_0, \beta_1, \dots, \beta_p) \in \mathbb{R}^{p+1}$, $\gamma = (\gamma_1, \dots, \gamma_q) \in \mathbb{R}^q$. Moreover, let Γ_q be the set of all $\gamma \in \mathbb{R}^q$ for which the lag polynomial $1 + \sum_{s=1}^q \gamma_s L^s$ has roots all out the unit circle. Under the conditions of theorem 10.3.3 (part (10.3.21)) there exists a subset N of $\mathbb{R}^{(p+1)k} \times \Gamma_q$ with Lebesgue measure zero such that $(\beta_{i,1}, \dots, \beta_{i,k}, \gamma_i) \in \mathbb{R}^{(p+1)k} \times \Gamma_q \setminus N$ implies

$$E(Z_{i,t} | Z_{t-1}, Z_{t-2}, \dots) = E(Z_{i,t} | \sum_{j=1}^k \psi(L|\beta_{i,j}, \gamma_i) Z_{j,t-1}) \text{ a.s.} \quad (10.3.26)$$

for $i=1, 2, \dots, k$ and $t=\dots, -2, -1, 0, 1, 2, 3, \dots$. Consequently, for each permissible rational lag polynomial $\psi(L|\beta_{i,j}, \gamma_i)$ there exist Borel measurable real function $f_{i,t}(\cdot)$ depending on $\beta_{i,1}, \beta_{i,2}, \dots, \beta_{i,k}$ and γ_i such that

$$E(Z_{i,t} | Z_{t-1}, Z_{t-2}, \dots) = f_{i,t}(\sum_{j=1}^k \psi(L|\beta_{i,j}, \gamma_i) Z_{j,t-1}) \text{ a.s.} \quad (10.3.27)$$

for $i=1, 2, \dots, k$ and $t=\dots, -2, -1, 0, 1, 2, 3, \dots$.

10.3.5 The nature of the ARMA memory index parameters and the response functions

In discussing the nature of the ARMA memory index parameters we shall first focus on the univariate case. Thus we now ask the question what the nature of a permissible τ in (10.3.1) is, i.e., is τ in general irrational or are also rational τ 's permissible? We recall that a permissible τ is such that the polynomial

$$\sum_{j=1}^m w_j \tau^{j-1} \quad (10.3.28)$$

is non-zero for arbitrary $m \geq 1$ and arbitrary rational numbers w_j not all equal to zero. But for $m=2$, $w_1 + w_2 \tau = 0$ for $\tau = -w_1/w_2$, so that for given rational τ we can always find rational numbers w_j such that the polynomial (10.3.28) equals zero. Obviously the same applies if only integer-valued w_j 's

are allowed. Hence the permissible τ 's are in general irrational. By a similar argument it can be shown that in the case of (10.3.26) at least some of the parameters in $\beta_{i,j}$ and γ_i will likely be irrational.

What is the consequence of the irrationality of the ARMA memory index parameters for the nature of the response functions? If we would pick an arbitrary permissible τ the Borel measurable real function $f_{t,\tau}$ for which

$$\begin{aligned} E(Z_t | Z_{t-1}, Z_{t-2}, \dots) &= E(Z_t | \sum_{j=1}^{\infty} \tau^{j-1} Z_{t-j}) \\ &= f_{t,\tau}(\sum_{j=1}^{\infty} \tau^{j-1} Z_{t-j}) \text{ a.s.} \end{aligned}$$

will likely be highly discontinuous, as this function has to sort out $Z_{t-1}, Z_{t-2}, Z_{t-3}, \dots$ from $\sum_{j=1}^{\infty} \tau^{j-1} Z_{t-j}$. See Sims(1988). On the other hand, if we choose τ such that

$$Z_t \text{ and } \sum_{j=1}^{\infty} \tau^{j-1} Z_{t-j}$$

are strongly correlated, and thus that

$$f_{t,\tau}(\sum_{j=1}^{\infty} \tau^{j-1} Z_{t-j}) \text{ and } \sum_{j=1}^{\infty} \tau^{j-1} Z_{t-j}$$

are strongly correlated, then the function $f_{t,\tau}$ will be close to a linear function. In any event, lemma 9.3.3 shows that the possible discontinuity of $f_{t,\tau}$ is not too dramatic, as $f_{t,\tau}$ can always be approximated arbitrarily close by a uniformly continuous function. Thus, given an arbitrary $\delta \in (0,1)$, a permissible $\tau \in (-1,1)$ and the condition $E|Z_t| < \infty$, there exists a uniformly continuous real function $g_{t,\tau}$ such that

$$E|f_{t,\tau}(\sum_{j=1}^{\infty} \tau^{j-1} Z_{t-j}) - g_{t,\tau}(\sum_{j=1}^{\infty} \tau^{j-1} Z_{t-j})| < \delta^2 \quad (10.3.29)$$

and consequently by Chebishev's inequality,

$$P(|f_{t,\tau}(\sum_{j=1}^{\infty} \tau^{j-1} Z_{t-j}) - g_{t,\tau}(\sum_{j=1}^{\infty} \tau^{j-1} Z_{t-j})| < \delta) \geq 1 - \delta. \quad (10.3.30)$$

We have argued that the ARMA memory index parameters are likely irrational. Since computers deal only with rational numbers we therefore cannot calculate the ARMA memory index exactly in practice. However, it is possible to choose a

rational τ_* close to τ such that "almost" all information about the past of the data generating process is preserved. The argument goes as follows. Due to the uniform continuity of $g_{t,\tau}$ there exist real numbers $\eta > 0$, $\rho \in (0,1)$ and a rational number τ_* close to τ such that

$$\begin{aligned} P(|g_{t,\tau}(\sum_{j=1}^{\infty} \tau^{j-1} Z_{t-j}) - g_{t,\tau}(\sum_{j=1}^{\infty} \tau_*^{j-1} Z_{t-j})| < \delta) \\ \geq P(|\sum_{j=1}^{\infty} \tau^{j-1} Z_{t-j} - \sum_{j=1}^{\infty} \tau_*^{j-1} Z_{t-j}| < \eta) \\ \geq P(|\tau - \tau_*| \sum_{j=2}^{\infty} (j-1) \rho^{j-2} |Z_{t-j}| < \eta), \end{aligned} \quad (10.3.31)$$

where $1 > \rho > \max(|\tau|, |\tau_*|)$. The last inequality follows from the mean value theorem. Thus by Chebishev's inequality

$$\begin{aligned} P(|g_{t,\tau}(\sum_{j=1}^{\infty} \tau^{j-1} Z_{t-j}) - g_{t,\tau}(\sum_{j=1}^{\infty} \tau_*^{j-1} Z_{t-j})| < \delta) \\ \geq 1 - |\tau - \tau_*| \sum_{j=2}^{\infty} (j-1) \rho^{j-2} E|Z_{t-j}| / \eta \geq 1 - \delta \end{aligned} \quad (10.3.32)$$

if

$$|\tau - \tau_*| \leq \delta \eta (1 - \rho)^2 / \sup_t E|Z_t|. \quad (10.3.33)$$

Combining (10.3.30) and (10.3.32) we see that for arbitrary $\delta \in (0, \frac{1}{2})$,

$$\begin{aligned} P(|f_{t,\tau}(\sum_{j=1}^{\infty} \tau^{j-1} Z_{t-j}) - g_{t,\tau}(\sum_{j=1}^{\infty} \tau_*^{j-1} Z_{t-j})| < \delta) \\ \geq 1 - 2\delta \text{ if (10.3.33) holds.} \end{aligned} \quad (10.3.34)$$

It should be noted that the rational-valued τ_* depends in general on the time index t . However, if the process (Z_t) is strictly stationary we can pick a constant τ_* , as is not too hard to verify from (10.3.29) through (10.3.34). In that case the functions $f_{t,\tau}$ and $g_{t,\tau}$ are independent of t . Summarizing, we have shown:

Theorem 10.3.5. Let (Z_t) be a strictly stationary univariate rational-valued process. Let assumptions 10.3.1 and 10.3.2 hold and let $\delta \in (0, \frac{1}{2})$ and $\tau \in (-1, 1) \setminus S$ be arbitrary, where S is the same as in theorem 10.3.1. There exists a uniformly continuous real function g_τ and a rational number τ_* in a neighborhood of

τ such that

$$P(|E(Z_t | Z_{t-1}, Z_{t-2}, \dots) - g_\tau(\sum_{j=1}^{\infty} \tau^{j-1} Z_{t-j})| < \delta) \geq 1-2\delta \quad (10.3.35)$$

Finally, a similar result can be shown for the general case in theorem 10.3.4. Thus:

Theorem 10.3.6. Let the conditions of theorem 10.3.4 hold and assume in addition that (Z_t) is strictly stationary. For arbitrary $\delta \in (0, \frac{1}{2})$ and $(\beta_{i,1}, \dots, \beta_{i,k}, \gamma_i) \in R^{(p+1)k} \times \Gamma_q \setminus N$ there exist uniformly continuous real functions g_i and vectors $(\beta_{i,1}^*, \dots, \beta_{i,k}^*, \gamma_i^*) \in Q^{(p+1)k} \times (\Gamma_q \cap Q^q)$ such that

$$P(|E(Z_{i,t} | Z_{t-1}, Z_{t-2}, \dots) - g_i(\sum_{j=1}^k \psi(L|\beta_{i,j}^*, \gamma_i^*) Z_{j,t-1})| < \delta) \geq 1-2\delta \quad (10.3.36)$$

for $i=1, 2, \dots, k$ and $t=\dots, -2, -1, 0, 1, 2, 3, \dots$

10.3.6 Discussion

We have shown that in modeling rational-valued time series processes as conditional expectations relative to the entire past of the process involved, it is possible to capture the relevant information about the past of the process by a single random variable, called an ARMA memory index. Given this ARMA memory index, the specification of the model then amounts to specifying a nonlinear response function defined on the real line. Although this response function might be highly discontinuous, it can be approximated arbitrarily close by a uniformly continuous real function of an ARMA memory index with rational-valued parameters.

One might argue that our approach is merely a sophisticated variant of representing a one-sided infinite sequence of variables as a decimal expansion of a real variable. For example, let the univariate stochastic process (Z_t) be integer-valued with values $0, 1, \dots, 9$, and define

$$\xi_t = \sum_{j=1}^{\infty} (0.1)^{j-1} Z_{t-j}.$$

Then ξ_t contains all the information about the past Z_{t-1}, Z_{t-2}, \dots of the process under review, hence

$$E(Z_t | Z_{t-1}, Z_{t-2}, \dots) = E(Z_t | \xi_t) \text{ a.s.}$$

In particular if

$$E(Z_t | Z_{t-1}, Z_{t-2}, \dots) = \psi(Z_{t-1}, Z_{t-2}, \dots),$$

where ψ is a real function on the space of one-sided infinite sequences of integers, then

$$\begin{aligned} E(Z_t | Z_{t-1}, Z_{t-2}, \dots) \\ = \psi([\xi_t], [10\xi_t] - 10[\xi_t], [10^2\xi_t] - 10[10\xi_t], [10^3\xi_t] - 10[10^2\xi_t], \dots) \\ = \psi^*(\xi_t), \end{aligned}$$

say, where $[x]$ denote truncation to the largest integer less or equal to x . Even if the function ψ is well-shaped, the function ψ^* is highly discontinuous. Moreover, one might argue that knowing ψ^* is not of much help, as it is impossible to store ξ_t exactly in the memory of a computer. Admittedly, the above primitive index is in general of limited use. The main contribution of our approach, however, is that a one-period ahead forecast on the basis of an almost arbitrary ARMAX model will work too. Modeling time series by ARMAX models can therefore be interpreted as specifying the response function $f_{1,t}$ in theorem 10.3.4 as a linear time invariant function, and estimation of an ARMAX model can be interpreted as looking for an ARMA memory index for which the response function is linear. Fitting an ARMAX model to the data forces the nonlinear response function towards a linear function. A good forecasting performance of the estimated ARMAX model then indicates that the corresponding response function $f_{1,t}$ is close to a linear function, as the best forecasting scheme is the one which represents the expectation of the dependent variable conditional on an ARMA memory index. Thus if one accepts ARMAX models as useful approximations of time series processes then actually one accepts the existence of a tractable ARMA memory index with corresponding response function close to a linear function.

The problem of storage of the ARMA memory index is not typical for our approach but a universal problem. For example, transforming the data by say a log transformation will result in loss of information, due to the finiteness of data storage in a computer. Whether this problem is serious or not for our ARMA memory index depends on the dependence of the data. Take for example the above primitive index ξ_t . Storing ξ_t as a double precision variable yields 29 significant decimal digits (in CDC Fortran5). Thus at least we can sort out Z_{t-1}, \dots, Z_{t-29} from ξ_t . If Z_t is almost independent of Z_{t-j} for $j > 29$ then

$$E(Z_t | Z_{t-1}, Z_{t-2}, \dots) \approx E(Z_t | Z_{t-1}, \dots, Z_{t-29}) \approx E(Z_t | \xi_t).$$

10.4 Nonlinear ARMAX models

The lesson we learn from the argument in the sections 10.2 and 10.3 is that the class of linear ARMAX models forms a good starting point for modeling vector time series processes. In modeling the conditional expectation

$$E(Z_{1,t} | Z_{t-1}, Z_{t-2}, \dots)$$

one should first look for the best fitting linear ARMAX model, as this strategy forces the nonlinear function f_1 , which maps the corresponding ARMA memory index $\xi_{1,t}$ into this conditional expectation, towards a linear function. Then apply various model misspecification tests to check the validity of the linear ARMAX model. We will consider these tests in the next chapter. If these tests indicate the presents of misspecification one could then try to model the nonlinear function f_1 , for which $E(Z_{1,t} | \xi_{1,t}) = f_1(\xi_{1,t})$, for example by specifying f_1 as a polynomial of a bounded one-to-one transformation of $\xi_{1,t}$, similarly to the approach in chapter 8. Moreover, one could run a nonparametric regression of $Z_{1,t}$ on $\xi_{1,t}$ to find a suitable functional form of f_1 . The latter approach is suggested in Bierens (1988a, section 6.2) and worked out further in Bierens (1988c). Also, plotting $Z_{1,t}$ and $\xi_{1,t}$ may reveal the form of this function f_1 . Thus, if the linear ARMAX model fails to pass model misspecification tests we may think of specifying a parametric family for the function f_1 , say $f(\cdot, \alpha)$, where α is a parameter vector. This approach gives rise to a model of the form

$$Y_t = f[(1 + \sum_{j=1}^q \gamma_j L)^{-1} (\sum_{j=1}^p \beta_j' Z_{t-j}), \alpha] + U_t.$$

where Y_t is one of the components of Z_t , (U_t) is the error process (which should now satisfy $E(U_t | Z_{t-1}, Z_{t-2}, \dots) = 0$ a.s.) and the β_j 's and $\gamma = (\gamma_1, \dots, \gamma_q)'$ are parameter vectors. In the sequel, however, we shall not deal with this class of models, for the simple reason that these models have not yet been considered in the literature, hence the sampling theory involved is yet absent. The main reason for introducing the ARMA memory index modeling theory is that it plays a key-role in our consistent model misspecification testing approach, in chapter 11.

Alternatively, if a linear ARMAX model does not pass our model misspecification tests one could follow Sims' (1988) common sense approach and add nonlinear terms to the best linear ARMAX model to capture the possible nonlinearity of the conditional expectation function. How these nonlinear terms should be specified depends on prior knowledge about the phenomena one wishes to model. This specification issue falls outside the scope of this book. Quoting Sims (1988): *There is no more general procedure available for inference in infinite-dimensional parameter spaces than the common sense one of guessing a set of finite-dimensional models, fitting them, and weighting together the results according to how well the models fit. This describes the actual behavior of most researchers and decision makers. (.....) Thus, if we are unsure of lag length and also believe that there may be nonlinearity in the system, a reasonable way to proceed is to introduce both a flexible distributed lag specification and some nonlinear terms that can be thought of as part of a Taylor or Fourier expansion of the nonlinear function to be estimated.* Sims' common sense approach will lead to a nonlinear ARMAX model of the form

$$Y_t = g(Z_{t-1}, \dots, Z_{t-p}, \beta) + U_t + \sum_{j=1}^q \gamma_j U_{t-j}. \quad (10.4.1)$$

where $g(\cdot, \beta)$ is a known parametric functional form for the AR part of the ARMAX model, with β a parameter vector. The MA part of this model may be considered as a flexible distributed lag specification, together with the AR lag structure implied by the function $g(\cdot, \beta)$.

In chapter 11 we consider the problem of estimating the parameters of model (10.4.1), taking the function $g(\cdot, \beta)$ as

given, and we derive the asymptotic properties of these estimators under strict stationarity of the data generating process (Z_t) as well as under data heterogeneity. Also, we consider various model misspecification tests, in particular consistent tests based on the ARMA memory index approach.

Remark: Admittedly, many important issues in time series analysis have not been discussed in this chapter. To mention a few, we have not paid attention to seasonal adjustment, unit roots and co-integration. This certainly does not mean that these issues are not important, but merely that they fall outside the scope of this book. As far as season and unit roots are concerned, it will be implicitly assumed that they have been removed by appropriate (seasonal) differencing. Moreover, we note that unit roots in time series can be detected by Phillips' (1987) version of Dickey and Fuller's (1979, 1981) tests. For co-integration we refer to Engle and Granger (1987).

References:

- Anderson, T.W. (1971), *The Statistical Analysis of Time Series*, New York: John Wiley
- Bierens, H.J. (1988a), "ARMA Memory Index Modeling of Economic Time Series (with discussion)", *Econometric Theory* 4, 1988, 35-59
- Bierens, H.J. (1988b), "Reply", *Econometric Theory* 4, 1988, 70-76
- Bierens, H.J. (1988c), "Model-Free Asymptotically Best Forecasting of Stationary Economic Time Series", RM1986-32, Faculty of Economics, Free University (revised: 1988)
- Dickey, D.A and W.A. Fuller (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", *Journal of the American Statistical Association* 74, 427-431
- Dickey, D.A and W.A. Fuller (1981), "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root", *Econometrica* 49, 1057-1072
- Engle, R.F. and C.W.J. Granger (1987), "Co-Integration and Error Correction: Representation, Estimation and Testing", *Econometrica* 55, 251-276

Phillips, P.C.B. (1987), "Time Series Regression with a Unit Root", *Econometrica* 55, 277-301

Royden, H.L. (1968), *Real Analysis*, London: Macmillan

Sargent, T.J. and C.A. Sims (1977), "Business Cycle Modeling Without Pretending to Have Too Much A Priori Economic Theory", in: C.A. Sims (ed.), *New Methods in Business Cycle Research: Proceedings from a Conference*, Minneapolis: Federal Reserve Bank of Minneapolis

Sims, C.A. (1981), "An autoregressive Index Model for the U.S., 1948-1975", in: J. Kmenta and J.B. Ramsey (ed.), *Large-Scale Macro-Econometric Models*, Amsterdam: North-Holland

Sims, C.A. (1988), "Comment on ARMA Memory Index Modeling of Economic Time Series", *Econometric Theory* 4, 64-69